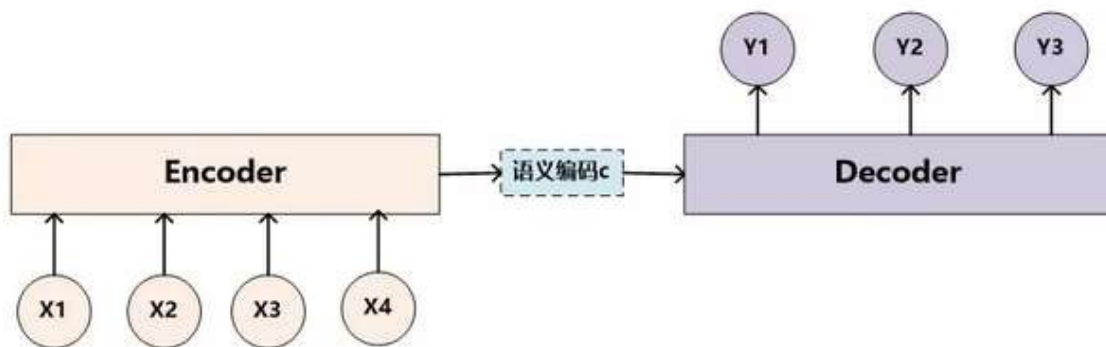


Encoder-Decoder 框架

Encoder-Decoder 框架可以看作是一种文本处理领域的研究模式，应用场景异常广泛，不仅仅可以用在对话机器人领域，还可以应用在机器翻译、文本摘要、句法分析等各种场合。下图是文本处理领域里常用的 Encoder-Decoder 框架最抽象的一种表示：



Encoder-Decoder框架

Encoder-Decoder 框架可以如此直观地去理解：可以把它看作适合处理由一个句子（或篇章）生成另外一个句子（或篇章）的通用处理模型。对于句子对 $\langle X, Y \rangle$ ，我们的目标是给定输入句子 X ，期待通过 Encoder-Decoder 框架来生成目标句子 Y 。X 和 Y 可以是同一种语言，也可以是两种不同的语言。而 X 和 Y 分别由各自的单词序列构成：

$$X = \langle x_1, x_2 \dots x_m \rangle$$
$$Y = \langle y_1, y_2 \dots y_n \rangle$$

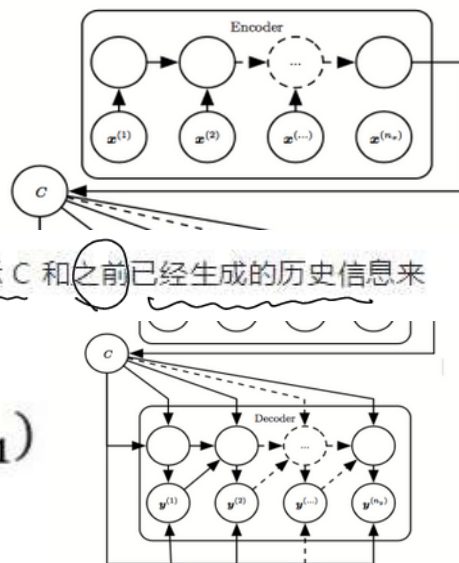
Encoder 顾名思义就是对输入句子 X 进行编码，将输入句子通过非线性变换转化为中间语义表示 C ：

$$C = \mathcal{F}(x_1, x_2 \dots x_m)$$

对于解码器 Decoder 来说，其任务是根据句子 X 的中间语义表示 C 和之前已经生成的历史信息来生成 i 时刻要生成的单词 y_i ：

$$y_i = \mathcal{G}(C, y_1, y_2 \dots y_{i-1})$$

中间语义



每个 y_i 都依次这么产生，那么看起来就是整个系统根据输入句子 X 生成了目标句子 Y 。

对于聊天机器人来说，完全可以使用上述的 Encoder-Decoder 框架来解决技术问题。具体而言，对应的 $\langle X, Y \rangle$ 中， X 指的是用户输入语句，一般称作 Message，而 Y 一般指的是聊天机器人的应答语句，一般称作 Response。其含义是当用户输入 Message 后，经过 Encoder-Decoder 框架计算，首先由 Encoder 对 Message 进行语义编码，形成中间语义表示 C ，Decoder 根据中间语义表示 C 生成了聊天机器人的应答 Response。这样，用户反复输入不同的 Message，聊天机器人每次都形成新的应答 Response，形成了一个实际的对话系统。

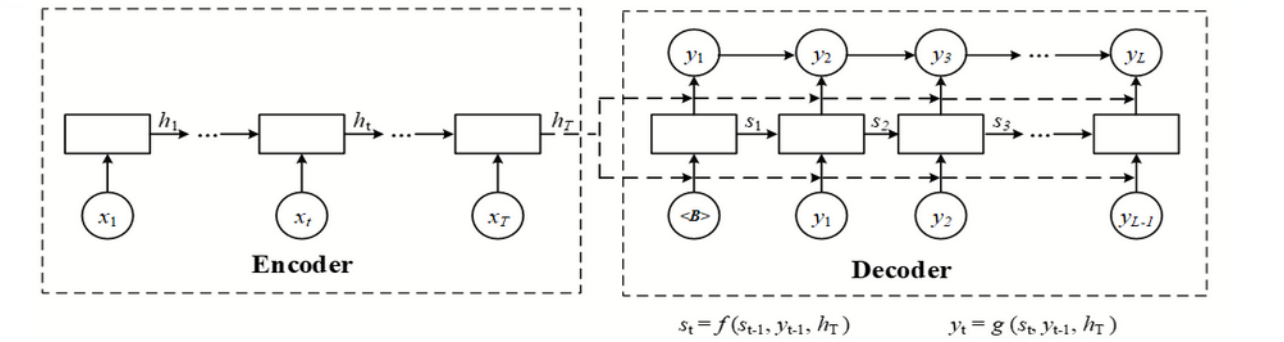
在实际实现聊天系统的时候，一般 Encoder 和 Decoder 都采用 RNN 模型，RNN 模型对于文本这种线性序列来说是最常用的深度学习模型，RNN 的改进模型 LSTM 以及 GRU 模型也是经常使用的模型，对于句子比较长的情形，LSTM 和 GRU 模型效果要明显优于 RNN 模型。尽管如此，当句子长度超过 30 以后，LSTM 模型的效果会急剧下降，一般此时会引入 Attention 模型，这是一种体现输出 Y 和输入 X 句子单词之间对齐概率的神经网络模型，对于长句子来说能够明显提升系统效果。

使用 Encoder-Decoder 模型来建立对话机器人的一般做法是：采用收集 Twitter 或者微博中评论里的聊天信息来作为训练数据，用大量的此类聊天信息来训练 Encoder-Decoder 模型中 RNN 对应的神经网络连接参数。下图展示了利用微博评论对话数据训练好的聊天机器人的聊天效果，其中 Post 列指的是用户输入 Message，其后三列是不同 Encoder-Decoder 方法产生的应答 Response，而最后一列是基于传统检索方法的输出应答。

ID	Post	NRM-glo	NRM-loc	NRM-hyb	Rtr-based
1	每年元旦都来高烧。 High fever attacks me every New Year's Day	多休息，增强免疫力！ Rest more to boost your immune system.	每年元旦都去医院看病 Go to see a doctor every New Year's Day	祝早日康复，身体健康！ Get well soon and stay healthy!	高烧... High fever...
2	又老了一岁，感谢我们组，我很高兴 I gain one more year. Grateful to my group, so happy	生日快乐！永远支持你！ Happy birthday! Will stand by you forever!	生日快乐！永远年轻漂亮！ Happy birthday! Wish you always stay young and pretty!	老了老了，岁月不饶人啊！ Getting old now. Time has no mercy.	我只是老了一岁 I just gain only one more year.
3	吃素第一天，坚持住，崔朵拉。 First day of being a vegetarian. Hold on, CuiDuoLa	必须来支持！加油！ I come to offer my support. Keep it up!	坚持住，坚持下去。 Hold on, keep it up.	每天都是素食主义者 I have been being a vegetarian everyday	坚持住！ Hold on!

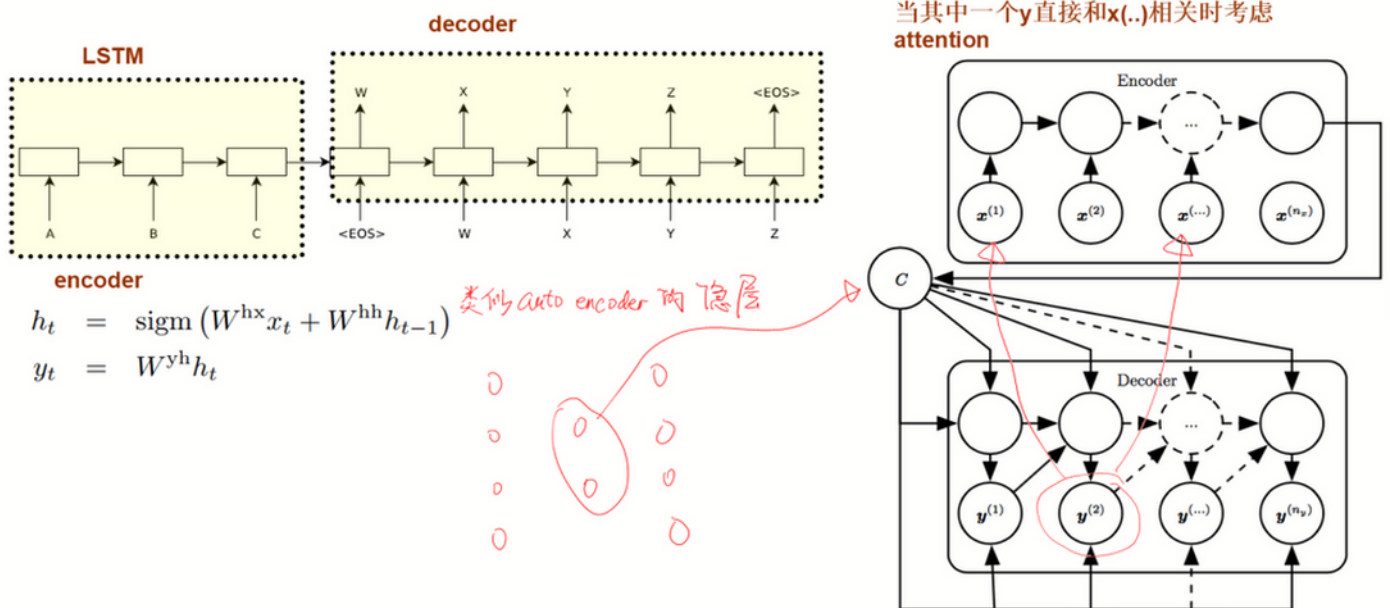
聊天机器人聊天效果示例

以上只是对 Encoder-Decoder 框架技术原理进行简单描述，聊天机器人的开发还需要考虑各种实际问题及其对应的解决方案，采用深度学习来进行聊天机器人的技术研发仍处于非常初期的阶段。



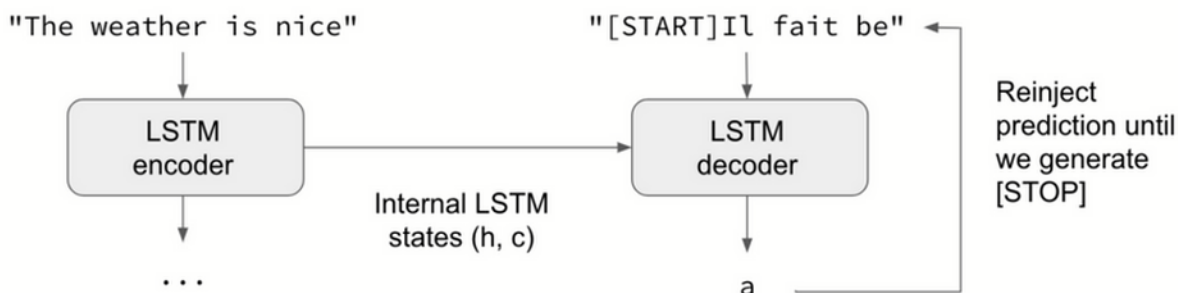
Seq2Seq

Bi-directional Lstm



supervise learning

Seq2Seq Example



- 1) Encode the input sequence into state vectors.
- 2) Start with a target sequence of size 1 (just the start-of-sequence character).
- 3) Feed the state vectors and 1-char target sequence to the decoder to produce predictions for the next character.
- 4) Sample the next character using these predictions (we simply use argmax).
- 5) Append the sampled character to the target sequence
- 6) Repeat until we generate the end-of-sequence character or we hit the character limit.